Vol. 160, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1989 Pages 79-86

HUMAN PROSTATIC ACID PHOSPHATASE: cDNA CLONING, GENE MAPPING AND PROTEIN SEQUENCE HOMOLOGY WITH LYSOSOMAL ACID PHOSPHATASE

Farida S. Sharief[1], Hansoo Lee[2], Mary M. Leuderman[3], Ake Lundwall[4], Larry L. Deaven[3], Ching-li Lee[2] and Steven S.-L. Li[1,5]

[1]Laboratory of Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709
[2]Department of Diagnostic Immunology Research and Biochemistry, Roswell Park Memorial Institute, Buffalo, New York 14263
[3]Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[4]Department of Clinical Chemistry, University of Lund, Malmo General Hospital, S-214 01 Malmo, Sweden

Received February 28, 1989

The cDNAs encoding human prostatic acid phosphatase were cloned and characterized. The mRNAs contain 3' noncoding regions of heterogeneous sizes 646, 1887 or 1913 nucleotides. A dimer and a monomer of the conserved Alu-repeats are present in the longer 3' noncoding sequences. The complete sequence of 354 amino acids for the mature enzyme was determined by sequencing both cDNA and protein. Human prostatic and lysosomal acid phosphatases exhibit 50% sequence homology, including five Cys residues and two putative N-linked glycosylation sites. The Acp-3 gene coding for human prostatic acid phosphatase was mapped onto chromosome 3 in this investigation. The Acp-2 gene coding for lysosomal acid phosphatase has previously been located on chromosome 11, while the Acp-1 gene coding for red blood cell acid phosphatase is on chromosome 2.

Acid phosphatases (Acp; EC 3.1.3.2) are a group of enzymes capable of hydrolysing phosphomonoesters under acidic conditions, and can be differentiated according to their immunological properties, tissue distribution and subcellular location(1). Since Gutman et al. (1936) first reported an elevation of serum acid phosphatase activity in patients with metastatic prostate cancer (2), human prostatic acid phosphatase (PAP) has been used as a diagnostic marker for prostate cancer (3). The physiological function of PAP is not well understood, but recent reports suggested that PAP

[5]To whom all correspondence should be addressed.

Abbreviations: Acp, Acid phosphatase; PAP, prostatic acid phosphatase; LAP, lysosomal acid phosphatase.

0006-291X/89 $1.50

Vol. 160, No. 1, 1989 B

may function in vivo as a phosphotyrosyl-protein phosphatase (4). The PAP is synthesized under androgen regulation by epithelial cells of the prostate, and it is secreted into the seminal fluid (5). In order to study the protein structure-function and mechanism of androgen regulation, we have undertaken the cDNA cloning and protein sequencing of human PAP. In this report, we describe the cDNA cloning and gene mapping of human PAP as well as protein sequence homology between human prostatic and lysosomal acid phosphatases.

Materials and Methods

Protein sequencing:

Human PAP protein was purified (6) and cleaved with CNBr and/or Lys-C protease (7). The resulting peptides were separated by HPLC microbore RP-C8 column using a linear gradient of acetonitrile (0-70% in 60 min) in 0.1% trifluoroacetic acid (8). The amino-terminal sequences of the mature enzyme and purified peptides were determined by automatic micro-sequencer equipped with on-line HPLC and data module (Applied Biosystems, Inc.).

cDNA cloning and sequencing:

A human prostate λgt11 cDNA expression library established from benign prostatic hyperplasia (9) was screened with rabbit antiserum against human PAP as previously described (10-12). The putative positive clones were purified to homogeneity using polyclonal and monoclonal antibodies against human PAP. The DNAs purified from the positive clones were analyzed by restriction endonuclease mapping and Southern blotting (13). The DNA fragments derived from EcoRI cleavage, as well as overlapping deletions constructed using an erase-a-base kit (Promega), were further subcloned into M13 mp18 phage (14), and the nucleotide sequences of the inserted DNAs were determined by the dideoxy chain termination method with the sequencing protocol modified to use deoxyadenosine 5-(α-35S)thiotriphosphate (15).

Gene mapping:

Human chromosomes were isolated and flow-sorted into 24 spots on nitrocellulose discs as previously described (16). These spot-blots and human genomic DNA blot were hybridized with 32P-labeled PAP cDNA probe as indicated in Figure 1. The conditions for hybridization and washing have been described previously (13,17).

Results

Protein sequencing:

The amino-terminal sequence of 45 residues from the purified PAP protein was obtained by micro-sequencing. A total of 80 amino acids (residue nos. 44-65, 148-153, 203-219, 240-251, 254-265 and 298-308) was also sequenced from four CNBr peptides and two Lys-C peptides (Fig. 2A). The amino acid sequence of one CNBr peptide was found to overlap with the amino-terminal sequence of the PAP protein. Thus, the sequence of the first 65 amino acids from mature PAP enzyme was unambigously established.

cDNA cloning and sequencing:

Six putative PAP cDNA clones were plaque-purified using polyclonal and monoclonal antibodies against PAP, and their inserted DNAs were shown to

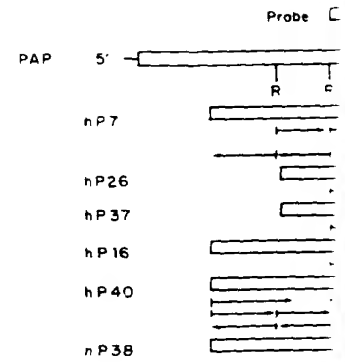

Fig. 1. Restriction enz
cDNA clones.
The restriction
those used in nucleotid
The direction and length
protein-coding sequence
Alu-repeats are shown by

cross-hybridize with each
of these six clones were
sequences determined as
sequence of these PAP cD
amino acids Phe-Leu-Asn
sequence, were identical
sequence of the purified
five peptides also conf
148-153, 203-219, 240-251
enzyme consists of a seq
did not possess the se
EcoRI-PstI cDNA fragment
used as a probe to iso
additional 24 PAP cDNA cl
codon no. 60.

As to the 3' noncoc
contained a poly (A)-tai
putative polyadenylation
this poly (A) addition
identical cDNA sequence
Clone hP16 did not posse
1442. Clone hP40 contair
no. 1887, while clone
nucleotide no. 1913.

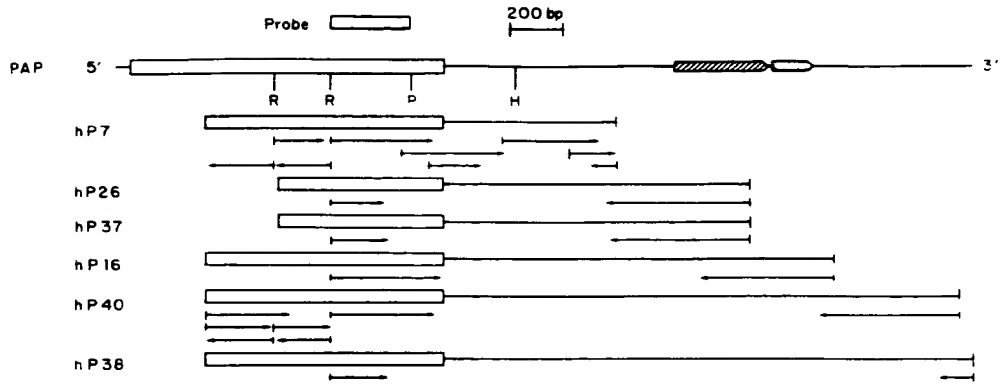

Fig. 1. Restriction endonuclease map and sequencing strategy of human PAP cDNA clones.
The restriction sites given (R, EcoRI; P, PstI and H, HindIII) are those used in nucleotide sequencing and preparation of hybridization probe. The direction and length of each sequencing run are indicated by arrows. The protein-coding sequence is denoted by open box. The dimeric and monomeric Alu-repeats are shown by hatched boxes.

cross-hybridize with each other (data not presented). The EcoRI DNA fragments of these six clones were subcloned into M13 mp18 phages, and their nucleotide sequences determined as indicated in Figure 1. The combined nucleotide sequence of these PAP cDNAs contain an open reading frame. The first six amino acids Phe-Leu-Asn-Glu-Ser-Tyr, which were deduced from the cDNA sequence, were identical with residue nos. 60 to 65 from the amino-terminal sequence of the purified PAP protein (Fig. 2A). The amino acid sequences of five peptides also confirmed the cDNA-deduced sequences of residue nos. 148-153, 203-219, 240-251, 254-265 and 298-308. Therefore, the purified PAP enzyme consists of a sequence of 354 amino acids. Since these six PAP cDNAs did not possess the sequence coding for the first 59 amino acids, an EcoRI-PstI cDNA fragment of coding region (as indicated in Fig. 1) was further used as a probe to isolate full-length PAP cDNA. However, none of the additional 24 PAP cDNA clones were found to contain the coding sequence 5′ to codon no. 60.

As to the 3' noncoding sequence of the PAP cDNAs (Fig. 3B), clone hP7 contained a poly (A)-tail of 51 residues added at nucleotide no. 646, and a putative polyadenylation signal of AATTAA was found at 20 nucleotides 5′ to this poly (A) addition site. Clones hP26 and hP37 appeared to contain identical cDNA sequence ending at nucleotide no. 1126 without poly (A)-tail. Clone hP16 did not possess a poly (A)-tail and terminated at nucleotide no. 1442. Clone hP40 contained a poly (A)-tail of 20 residues added at nucleotide no. 1887, while clone hp38 had a poly (A)-tail of 41 residues added at nucleotide no. 1913. Both hP40 and hP38 clones contained a common

```
                                                               ggccagaaacagctctcctcaacatgaga
                                                                                  -32  MetArg
A    gctgcacccctcctcctggccagggcagcaagcttagccttggcttcttgtttctgctttttttgctggctagaccgaagtgtactagcc
     AlaAlaProLeuLeuLeuAlaArgAlaAlaSerLeuAlaLeuAlaSerCysPheCysPhePheCysTrpLeuAspArgSerValLeuAla

     aaggagttgaagtttgtgactttggtgtttcggcatggagaccgaagtcccattgacacctttcccactgaccccataaaggaatcctca
  1  LysGluLeuLysPheValThrLeuValPheArgHisGlyAspArgSerProIleAspThrPheProThrAspProIleLysGluSerSer

     tggccacaaggatttggccaactcacccagctgggcatggagcagcattatgaacttggagagtatataagaaagagatatagaaAATTC
 31  TrpProGlnGlyPheGlyGlnLeuThrGlnLeuGlyMetGluGlnHisTyrGluLeuGlyGluTyrIleArgLysArgTyrArgLysPhe

     TTGAATGACTCCTATAAACATGAACAGGTTTATATTCGAAGCACAGACGTTGACCGGACTTTGATGAGTCGTATGACAAACCTGGCAGCC
 61  LeuAsnGluSerTyrLysHisGluGlnValTyrIleArgSerThrAspValAspArgThrLeuMetSerAlaMetThrAsnLeuAlaAla

     CTGTTTCCCCCAGAAGGTGTCAGCATCTGGAATCCTATCCTACTCTGGCAGCCCATCCCGGTGCACACAGTTCCTCTTTCTGAAGATCAG
 91  LeuPheProProGluGlyValSerIleTrpAsnProIleLeuLeuTrpGlnProIleProValHisThrValProLeuSerGluAspGln

     TTGCTATACCTGCCTTTCAGGAACTGCCCTCGTTTTCAAGAACTTGAGAGTGAGACTTTGAAATCAGAGGAATTCCAGAAGAGGCTGCAC
121  LeuLeuTyrLeuProPheArgAsnCysProArgPheGlnGluLeuGluSerGluThrLeuLysSerGluGluPheGlnLysArgLeuHis

     CCTTATAAGGATTTTATAGCTACCTTGGGAAAACTTTCAGGATTACATGGCCAGGACCTTTTTGGAATTTGGAGTAAAGTCTACGACCCT
151  ProTyrLysAspPheIleAlaThrLeuGlyLysLeuSerGlyLeuHisGlyGlnAspLeuPheGlyIleTrpSerLysValTyrAspPro

     TTATATTCTGAGAGTGTTCACAATTTCACTTTACCCTCCTGGGCCACTGAGGACACCATGACTAAGTTGAGAGAATTGTCAGAATTGTCC
181  LeuTyrCysGluSerValHisAsnPheThrLeuProSerTrpAlaThrGluAspThrMetThrLysLeuArgGluLeuSerGluLeuSer

     CTCCTGTCCCTCTATGGAATTCACAAGCAGAAAGAGAAATCTAGGCTCCAAGGGGGTGTCCTGGTCAATGAAATCCTCAATCACATGAAG
211  LeuLeuSerLeuTryGlyIleHisLysGlnLysGluLysSerArgLeuGlnGlyGlyValLeuValAsnGluIleLeuAsnHisMetLys

     AGAGCAACTCAGATACCAAGCTACAAAAAACTTATCATGTATTCTGCGCATGACACTACTGTGACTGGCCTACAGATGGCGCTAGATGTT
241  ArgAlaThrGlnIleProSerTyrLysLysLeuIleMetTyrSerAlaHisAspThrThrValSerGlyLeuGlnMetAlaLeuAspVal

     TACAACGGACTCCTTCCTCCCTATGCTTCTTGCCACTTGACGGAATTGTACTTTGAGAAGGGGGAGTACTTTGTGGAGATGTACTACCGG
271  TyrAsnGlyLeuLeuProProTyrAlaSerCysHisLeuThrGluLeuTyrPheGluLysGlyGluTyrPheValGluMetTryTryArg

     AATGAGACGCAGCACGAGCCGTATCCCCTCATGCTACCTGGCTGCAGCCCCAGCTGTCCTCTGGAGAGGTTTGCTGAGCTGGTTGGCCCT
301  AsnGluThrGlnHisGluProTyrProLeuMetLeuProGlyCysSerProSerCysProLeuGluArgPheAlaGluLeuValGlyPro

     GTGATCCCTCAAGACTGGTCCACGGAGGTTATGACCACAAACAGCCATCAAGGTACTGAGGACAGTACAGATTAG
331  ValIleProGlnAspTrpSerThrGluValMetThrThrAsnSerHisGlnGlyThrGluAspSerThrAsp   354
```

```
             10        20        30        40        50        60        70        80        90       100
B    1  TGTGCACAGAGATCTCTGTAGAAAGAGTAGCTGCCCTTTCTCAGGGCAGATGATGCTTTGAGAACATACTTTGGCCATTACCCCCCAGCTTTGAGGAAAA
   101  TGGGCTTTGGATGATTATTTTATGTTTTAGGGGACCCCCAACCTCAGGCAATTCCATCCTCTTCACCCGACCCTGCCCCCACTTGCCATAAAACTTAGCT
   201  AAGTTTTGTTTTGTTTTTCAGCGTTAATGTAAAGGGGCAGCAGTGCCAAAATATAACAGAGATAAAGCTTAGGTCAAAGTTCATAGAGTTCCCATGAACT
   301  ATATGACTGGCCACACAGGATCTTTTGTATTTAAGGATTCTGAGATTTTGCTTGAGCAGGATTAGATAAGGCTGTTCTTTAAATGTCTGAAATGGAACAG
   401  ATTTCAAAAAAAACCCCACAATCTAGGGTGGGAACAAGGAAGGAAAGATGTGAATAGGCTGATGGGCAAAAAACCAATTTACCCATCAGTTCCAGCCTTC
   501  TCTCAAGGAGAGGCAAAGAAAGGAGATACAGTGGAGACATCTGGAAAGTTTTCTCCACTGGAAAACTGCTACTATCTGTTTTTATATTTCTGTTAAAATA
   601  TATGAGGCTACAGAACTAAAAATTAAAACCTCTTTGTGTCCCTTGGTCCTGGAACATTTATGTTCCTTTTAAAGAAACAAAAATCAAACTTTACAGAAAG
   701  ATTTGATGTATGTAATACATATAGCAGCTCTTGAAGTATATATATCATAGCAAATAAGTCATCTGATGAGAACAAGCTATTTGGGCACAACACATCAGGA
   801  AAGAGAGCACCACGTGATGGAGTTTCTCCAGAAGCTCCAGTGATAAGAGATGTTGACTCTAAAGTTGATTTAAGGCCAGGCATGGTGGTTTACGCCTATA
   901  ATCCCAGCATTTTGGGAGTCCGAGGTGGGCAGATCACTTGAGCTCAGGAGGTCAAGATCAGCCTGGGCAACATGGTGAAACCTTGTCTCTACATAAAATA
  1001  CAAAAACTTAGATGGGCATGGTGGTGTGTGCCTATAGTCCACTACTTGTGGGGCTAAGGCAGGAGGATCACTTGAGCCCCGGAGGTCGAGGCTACAGTGA
  1101  GCCAAGAGTGCACTACTGTACTCCAGCCAGGGCAAGAGAGCGAGACCCTGTCTCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA
  1201  TAAAAACAAAGTTGATTAAGAAAGGAAGTATAGGCTAGGCACAGTGGCTCACACCTGTAATCCTTGCATTTTGGAAGGCTGAGGCAGGAGGATCACTTTA
  1301  GGCCTGGTGTGTTCAAGACCAGCCTGGTCAACATAGTGAGACACTGTCTCTACCAAAAAAAGGAAGGAAGGGACACATATCAAACTGAAACAAAATTAGA
  1401  AATGTAATTATGTTATGTTCTAAGTGCCTCCAAGTTCAAAACTTATTGGAATGTTGAGAGTGTGGTTACGAAATACGTTAGGAGGACAAAAGGAATGTGT
  1501  AAGTCTTTAATGCCCGATATCTTCAGAAAACCTAAGCAAACTTACAGGTCCTGCTGAAACTGCCCACTCTGCAAGAAGAAATCATGATATAGCTTTGCCA
  1601  TGTGGCAGATCTACATGTCTAGAGAACACTGTGCTCTATTACCATTATGGATAAAGATGAGATGGTTTCTAGAGATGGTTTCTACTGGCTGCCAGAATCT
  1701  AGAGCAAAGCCATCCCCGCTCCTGGTTGGTCACAGAATGACTGACAAAGACATCGATTGATATGCTTCTTTGTGTTATTTCCCTCCCAAGTAAATGTTTG
  1801  TCCTTGGGTCCATTTTCTATGCTTGTAACTGTCTTCTAGCAGTGAGCCAAATGTAAAATAGTGAATAAAGTCATTATTAGGAAGTTCAAAAGCATTGCTT
  1901  TTATAATGAACTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 2. The nucleotide and amino acid sequences of human PAP.

A. The amino acid sequence of PAP is numbered from the first residue of the mature enzyme, and the number of the first amino acid in each row is given on the left. The amino acid sequences obtained by direct protein sequencing are underlined. The nucleotide sequence 5' to codon no. 60 is from the previously reported sequence (20) in which a sequence of Trp-Ile-Trp-Pro-Thr-His-Pro-Ala was deduced for amino acid nos. 34 to 41. Further, Ala and Cys were reported for amino acid nos. 180 and 340, respectively (20). It may also be noted that a Gln instead of Glu at amino acid no. 2 was detected previously (10).

B. The 3' noncoding sequence of human PAP is numbered from the first nucleotide after termination codon TAG, and the number of the first nucleotide in each row is given on the left. Clones hP7, hP40 and hP38 have poly (A) addition sites (indicated by dots) at nucleotide nos. 646, 1887 and 1913, respectively. The dimer and monomer of Alu-repeats are underlined. The previously reported sequence (20) and the sequence determined in this investigation exhibit five substitutions at nucleotide nos. 156 (T for A), 157 (A for T), 168 (T for C), 385 (T for G) and 1211 (C for G); a single nucleotide (T) insertion after residue no. 256 and a double nucleotide (TT) addition after residue no. 413; eight deletions at nucleotide nos. 417, 1206, 1210, 1219, 1223, 1227, 1243 and 1842.

R P

A

Fig. 3. Analyses of h
PAP cDNA prob
A. The to
endonucleases EcoRI (
0.75% agarose gel.
filter and hybridized
in Fig. 1. The estim
and PstI, and 3.8Kb fc
B. The humar
nitrocellulose discs
hybridized with the s

polyadenylation signal
respectively, upstream
hP37, hP16, hP40 and hF
3' noncoding sequence
sequences. The dimeri
by direct repeats of
monomeric Alu-sequence
nos. 1222-1370).

Gene mapping:

The genomic blot
showed a single hybric
of a strong band and
cleavage suggested the
probe, since this cD
HindIII site. The res

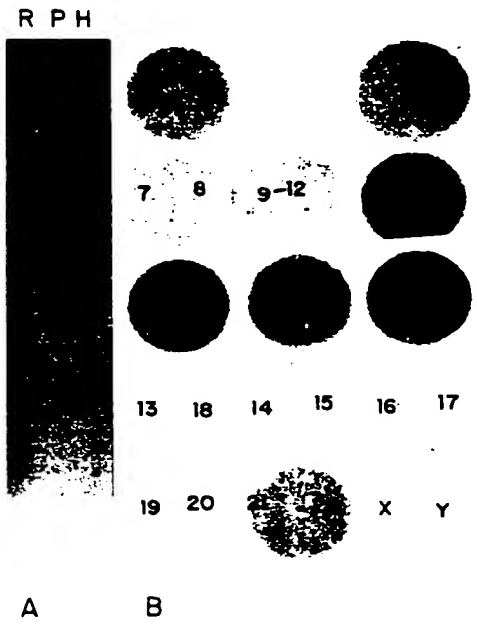

Fig. 3. Analyses of human genomic blot and chromosome spot-blots using human PAP cDNA probe.
A. The total human genomic DNA was cleaved with restriction endonucleases EcoRI (R), PstI (P) or HindIII (H), and electrophoresed on 0.75% agarose gel. The DNA fragments were transferred to nitrocellulose filter and hybridized with 32P-labeled coding probe of PAP cDNA as indicated in Fig. 1. The estimated sizes of major DNA fragments are: 4.6Kb for EcoRI and PstI, and 3.8Kb for HindIII.
B. The human chromosomes of each type were flow-sorted directly onto nitrocellulose discs as previously described (16). The spot-blots were hybridized with the same coding probe used in genomic blot analysis.

polyadenylation signal of AATAAA located at 18 and 44 nucleotides, respectively, upstream to their poly (A) addition sites. Thus, clones hP26, hP37, hP16, hP40 and hP38 appeared to have originated from mRNAs in which the 3′ noncoding sequences contain a dimer and a monomer of Alu-repetitive sequences. The dimeric Alu-sequence including 12 repeats of TAAA is flanked by direct repeats of AAAGTTGATT (nucleotide nos. 861-1216), while the monomeric Alu-sequence is surrounded by direct repeats of AAGGAAG (nucleotide nos. 1222-1370).

Gene mapping:

The genomic blot analysis of human total DNA using PAP coding probe showed a single hybridization band in PstI cut DNA (Fig. 3A). The detection of a strong band and a weak band in the DNAs of either EcoRI or HindIII cleavage suggested the presence of an intron within the cDNA region used as a probe, since this cDNA sequence contains neither an internal EcoRI nor a HindIII site. The results of chromosome spot-blot hybridization with the same

PAP coding probe indicated that only chromosome 3 exhibits a positive signal (Fig. 3B). Thus, the gene coding for human PAP is located on chromosome 3.

Discussion

The complete amino acid sequence of purified human PAP was determined by sequencing both protein and cDNA, and the mature enzyme contains a sequence of 354 amino acids, including 10 Ala, 15 Arg, 10 Asn, 16 Asp, 5 Cys, 17 Gln, 29 Glu, 19 Gly, 13 His, 14 Ile, 44 Leu, 19 Lys, 10 Met, 15 Phe, 25 Pro, 27 Ser, 24 Thr, 6 Trp, 19 Tyr and 17 Val. Three putative N-linked glycosylation sites of PAP were found at Asn-Glu-Ser (residues 62-64), Asn-Phe-Thr (residues 188-190) and Asn-Glu-Thr (residues 301-303). The glycosylation at these three potential sites of PAP may explain the difference between the estimated 50 kDa of glycosylated subunits (6) and the calculated Mr of 40,939 for mature polypeptide chains.

Human PAP mRNAs have heterogeneous sizes of 3′ noncoding region. Clone hP7 contain only 646 nucleotides of the 3′ noncoding sequence; while clones hP40 and hP38 have 3' noncoding regions of 1887 and 1913 nucleotides, respectively. A dimer and a monomer of Alu-repeats are present in the longer 3' noncoding sequences, and these two Alu-repeats exhibit approximately 80% homology with the consensus Alu-sequence (18). It is of interest that an Alu-repeat is also present in the 3' noncoding region of the cDNA encoding for human placental alkaline phosphatase (19). However, the biological significance of the Alu-repeats in the 3′ noncoding region of PAP mRNAs remains to be determined.

This investigation found that none of the isolated 30 PAP cDNA clones contain the coding sequence 5' to codon no. 60, while several PAP cDNAs truncated at the same AAATTC site of codon no. 59-60. This result may indicate that EcoRI restriction endonuclease cleaves the EcoRI-like sequence, but EcoRI methylase did not protect this sequence during the cDNA cloning (9). Thus, a different procedure without using EcoRI restriction endonuclease must be utilized to establish the cDNA library in order to obtain full-length PAP cDNA.

While this work was in progress, human PAP cDNA was independently cloned (20). A comparison of protein sequence revealed 10 amino acid differences at residue nos. 34 to 41, 180 and 340. The differences between the sequence of Gly-Phe-Gly-Gln-Leu-Thr-Gln-Leu obtained by direct protein sequencing in this investigation and the predicted sequence of Trp-Ile-Trp-Pro-Thr-His-Pro-Ala can simply be due to an insertion of T at codon no. 34 and deletion of G at codon no. 41 in the reported cDNA sequence (20). It should also be noted that the amino acid sequence of reside nos. 34 to 39 obtained by direct protein sequencing is identical to the corresponding sequence of human LAP (21) as

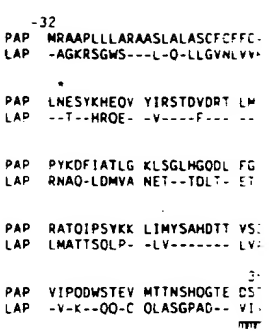

Fig. 4. Comparison of
The amino acid
identical residues are
addition/deletion in PA
shown by dots, while th
star. The transmembran

indicated in Figure 4. and 340 (Val vs Cys) may At the 3' noncoding regio substitutions and 11 adc and the sequence determir

The gene coding fo gene is designated as Ac phosphatase is located c acid phosphatase (LAP) is human LAP was reported t sequence (21). There amino acid sequences of as well as the first anc sequence were conserved proteins contain hydrop respectively (20-21). Bc L-tartrate (6,21). Thes the genes coding for hur to a multi-gene family o evolution.

Acknowledgments

We thank Drs. B. H. of this project, Mr. R. H Drs. F. Johnson and R. A Ms. L. Roberts for typing

```
     -32                             +1                              30                                 60
PAP  MRAAPLLLARAASLALASCFCFFCWLORSVLA KELKFVTLVF RHGDRSPIDT FPTKPIKESS WPOGFGQLTO LGMEOHYELG EYIRKRYRKF
LAP  -AGKRSGWS---L-Q-LLGVNLVVMPPTRΔΔ- RS-R----LY -------VK- Y-KD-YQ-EE ---------K E--L--W--- OAL-Q--HG-

      *                        90                                120          :                    150
PAP  LNESYKHEQV YIRSTDVDRT LMSAMTNLAA LFPPEGVSIW NPILLWQPIP VHTVPLSEDQ LLYLPFRNCP RFQELESETL KSEEFQKRLH
LAP  --T--HRQE- -V----F--- ----EA---G ----N-MQRF --NIS----- -----IT--R --KF-LGP-- -YEQ-QN--R QTP-Y-NESS

                               180  :    *                       210                                240
PAP  PYKDFIATLG KLSGLHGQDL FGIWSKVYDP LYCESVHNFT LPSWATEDTMT KLRELSELS LLSLYGIHKQ KEKSRLQGGV LVNEILNHMK
LAP  RNAQ-LDMVA NET--TDLT- ETV-NΔ---T -F--QT-GLR --P--SPQ--Q R-SR-KDF- FRF-F--YQ- A--A------ -LAQ-RKNLT

                               270           :                   300*             :   :            330
PAP  RATQIPSYKK LIMYSAHDTT VSGLQMALDV YNGLLPPYAS CHLTELYFEK GEYFΔVEMYYRNETQHEPYPL MLPGCSPSCP LERFAELVGP
LAP  LMATTSQLP- -LV------- LVA------- ---EQA---- --IF---Q-D SGN-S----F---SDKA-W-- S----PHR-- -QD-LRLTE-

                          354
PAP  VIPQDWSTEV MTTNSHQGTE DSTD
LAP  -V-K--QQ-C QLASGPAD-- VIVALAVCGSILFLLIVLLLTVLFRMQAOPPGYRHVADGEDHA
                           mmmmmmmmmmmmmmmmmmmmmmmmm            393
```

Fig. 4. Comparison of amino acid sequences of human PAP and LAP. The amino acids of LAP (21) different from PAP are given; while those identical residues are indicated by hyphen. The open triangle means the addition/deletion in PAP and LAP sequences. The conserved Cys residues are shown by dots, while the potential N-linked glycosylation sites are labeled by star. The transmembrane peptide of LAP is denoted by m.

indicated in Figure 4. The differences at amino acid nos. 180 (Pro vs Ala) and 340 (Val vs Cys) may represent genetic polymorphisms in human population. At the 3' noncoding region of 1913 nucleotides (Fig. 2B) there are 5 nucleotide substitutions and 11 additions/deletions between the reported sequence (20) and the sequence determined in this investigation.

The gene coding for PAP was mapped onto human chromosome 3, and this gene is designated as Acp-3. The Acp-1 gene coding for red blood cell acid phosphatase is located on chromosome 2; the Acp-2 gene coding for lysosomal acid phosphatase (LAP) is on the short arm of chromosome 11 (21-23). Recently, human LAP was reported to contain 393 amino acids, including a transmembrane sequence (21). There is 50% (177/351) homology between the corresponding amino acid sequences of human PAP and LAP (Fig. 4). All five Cys residues as well as the first and third potential N-linked glycosylation sites of PAP sequence were conserved in the LAP sequence. Human PAP and LAP precursor proteins contain hydrophobic signal peptides of 32 and 30 amino acids, respectively (20-21). Both PAP and LAP enzymes are sensitive to inhibition by L-tartrate (6,21). These structural and functional similarities suggest that the genes coding for human prostatic and lysosomal acid phosphatases belong to a multi-gene family originated from an ancestral gene during the course of evolution.

Acknowledgments

We thank Drs. B. H. Judd and T. M. Chu for their interest in and support of this project, Mr. R. Henry for his expert assistance in protein sequencing, Drs. F. Johnson and R. Allen for reading the manuscript, and Ms. N. Gore and Ms. L. Roberts for typing it.

References

1. Yam, L.T. (1974) Am. J. Med. 56, 604-616.
2. Gutman, E.B., Sproul, E.E., and Gutman, A.B. (1936) Am. J. Cancer 28, 485-495.
3. Chu, T.M., Wang, M.C., Lee, C.L., Killian, C.S., and Murphy, G.P. (1982) In Biochemical Markers for Cancer (Shu, T.M., ed.) pp 117-136, Marcel Dekker, New York.
4. Lin, M.F., and Clinton, G.M. (1987) Adv. Protein Phosphatases 4, 199-228.
5. Yam, L.T., Li, C.Y., and Lam, K.W. (1980) In Spring-Mills, E. and Hafez, E.S.E. (eds.) Male Accessory Sex Glands. Elsevier/North Holland. Amserstam, pp. 183-196.
6. Lin, M.F., Lee, C.L., Li, S.S.-L., and Chu, T.M. (1983) Biochem. 22, 1055-1061.
7. Pan, Y.-C.E., Sharief, F.S., Okabe, M., Huang, S., and Li, S.S.-L. (1983) J. Biol. Chem. 258, 7005-7016.
8. Hunkapiller, M.W., Hewick, R.M., Dreyer, W.J., and Hood, L.E. (1983) Methods Enzymol. 91, 399-413.
9. Lundwall, A., and Lilja, H. (1987) FEBS Letters 214, 317-322.
10. Lin, M.F., Lee, C.L., Sharief, F.S., Li, S.S.-L., and Chu, T.M. (1983) Cancer Res. 43, 3841-3846.
11. Young, R.A., and Davis, R.W. (1985) In Genetic Engineering (Setlow, J. and Hollaender, A., eds.) Vol. 7, pp. 29-41, Plenum, New York.
12. Sakai, I., Sharief, F.S., and Li, S.S.-L. (1987) Biochem. J. 242, 619-622.
13. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
14. Messing, J., Crea, R., and Seeberg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
15. Biggin, M.C., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
16. Deaven, L.L., van Dilla, M.A., Bartholdi, M.F., Carrano, A.V., Cram, L.S., Fuscoe, J.C., Gray, J.W., Hildebrand, C.E., Moyzis, R.K., and Perlman, J. (1986) Cold Spring Harb. Symp. Quant. Biol. 51, 159-167.
17. Li, S.S.-L., Luedemann, M., Sharief, F.S., Takano, T., and Deaven, L.L. (1988) Cytogenet. Cell Genet. 48, 16-18.
18. Jurka, J., and Smith, T. (1988) Proc. Natl. Acad. Sci. USA 85, 4775-4778.
19. Henthorn, P.S., Knoll, B.J., Raducha, M., Rothblum, K.N., Slaughter, C., Weiss, M., Lafferty, M.A., Fischer, T., and Harris, H. (1986) Proc. Natl. Acad. Sci. USA 83, 5597-5601.
20. Vihko, P., Virkkunen, P., Henttu, P., Roiko, K., Solin, T., and Huhtala, M.L. (1988) FEBS Letters 236, 275-281.
21. Pohlmann, R., Krentler, C., Schmidt, B., Schroder, W., Lorkowski, G., Culley, J., Mersmann, G., Geier, C., Waheed, A., Gottschalk, S., Grezschik, K.H., Hasilik, A., and von Figura, K. (1988) EMBO J. 7, 2343-2350.
22. Busby, N., Courval, J., and Francke, U. (1976) Cytogenet. Cell Genet. 16, 105-107.
23. Shows, T.B., Brown, J.A., and Lalley, P.A. (1976) Cytogenet. Cell Genet. 16, 231-234.